Research Article | Vol. 11, No. 6 / 1 June 2020 / Biomedical Optics Express 3095
Biomedical Optics EXPRESS

# Adversarial convolutional network for esophageal tissue segmentation on OCT images

CONG WANG,[1,3] MENG GAN,[1,3,*] MIAO ZHANG,[2] AND DEYIN LI[2]

[1] *Jiangsu Key Laboratory of Medical Optics, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China*
[2] *School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China*
[3] *These authors contributed equally to this work and should be considered co-first authors*
[*] *ganm@sibet.ac.cn*

**Abstract:** Automatic segmentation is important for esophageal OCT image processing, which is able to provide tissue characteristics such as shape and thickness for disease diagnosis. Existing automatical segmentation methods based on deep convolutional networks may not generate accurate segmentation results due to limited training set and various layer shapes. This study proposed a novel adversarial convolutional network (ACN) to segment esophageal OCT images using a convolutional network trained by adversarial learning. The proposed framework includes a generator and a discriminator, both with U-Net alike fully convolutional architecture. The discriminator is a hybrid network that discriminates whether the generated results are real and implements pixel classification at the same time. Leveraging on the adversarial training, the discriminator becomes more powerful. In addition, the adversarial loss is able to encode high order relationships of pixels, thus eliminating the requirements of post-processing. Experiments on segmenting esophageal OCT images from guinea pigs confirmed that the ACN outperforms several deep learning frameworks in pixel classification accuracy and improves the segmentation result. The potential clinical application of ACN for detecting eosinophilic esophagitis (EoE), an esophageal disease, is also presented in the experiment.



## 1. Introduction

Optical coherence tomography (OCT) is an imaging technique proposed by Huang, et al. [1] in 1991, which is able to image biological tissues in high resolution. It was first used in ophthalmology to help detect eye diseases [1]. In 1997, Tearney et al. combined OCT with the fiber-optic flexible endoscopes to enter the upper gastrointestinal tract [2]. Leveraging on Tearney's work, the OCT device can be used to image the microstructure of esophagus, which helps to diagnose a variety of esophageal diseases, such as Barrett's esophagus (BE) [3], eosinophilic esophagitis (EoE) [4], and dysplasia [5]. Disease diagnosis using OCT equipment is laborious since it relies on accurate interpretation of a large number of images, and a computer-assisted analysis system can help deal with this problem. Researchers have developed automatical systems for diagnosing esophageal diseases like BE by classifiers [6,7]. However, an explainable diagnosis based on tissue characteristics such as shape, thickness and area is more attractive since it is more intuitive and reliable than a "black box" classification system [8]. Tissue segmentation is the key technique in such an explainable disease diagnosis process.

Representative methods for automatical esophageal tissue layer segmentation can be summarized as follows. In 2016, Ughi et al. proposed an A-scan based method for esophageal lumen segmentation [9], but it can hardly be generalized for segmenting internal tissue layers. Then in 2017, Zhang et al. [10] employed the graph-based method to segment five clinical-related tissue layers, realizing multi-layer esophageal tissue segmentation. Inspired by Zhang's research, our group proposed an edge-enhanced graph search method to achieve more accurate esophageal OCT image segmentation [11]. The graph-based method requires a priori knowledge like tissue

#394715 https://doi.org/10.1364/BOE.394715
Journal © 2020 Received 8 Apr 2020; revised 8 May 2020; accepted 8 May 2020; published 18 May 2020

Biomedical Optics EXPRESS

width, which limits its application in some irregular cases. To solve this problem our group designed an automatical segmentation system based on wavelet features and sparse Bayesian classifier in 2019, which is more robust than the traditional gradient-based strategy [12]. Almost at the same time, Li et al. proposed a U-Net based framework for an end-to-end esophageal layer segmentation, which introduces deep learning algorithms to the community of esophageal OCT image processing [8].

In recent years, deep convolutional network is becoming the primary approach in computer vision tasks leveraging on its superior performance and easy implementation. Although the typical application of deep convolutional network is classification [13–15], many researchers attempted to use it to address problems of biomedical image segmentation [16,17]. In the community of OCT image segmentation, deep learning based strategies are also treated as the state-of-the-arts [18–20]. A commonly used idea is identifying tissue layer boundaries by classifying image patches using a deep neural network. For example, Fang et al. segmented nine tissue layers in retinal OCT images based on the patch classification result of a convolutional neural network [21]. Kugelman et al. identified the retinal boundaries using recurrent neural networks and graph search [22]. Although the segmentation results are promising, such methods usually suffer from large redundancy and result in more inference time [22]. A more elegant framework is pixel classifying by the fully convolutional network (FCN) [23,24]. This kind of method takes advantage of convolutional networks and uses an encoder-decoder architecture to assign each pixel to a label. A most widely employed work is proposed by Ronneberger, which designed a U-shape FCN called U-Net to deal with biomedical images with small training set [25]. Based on FCN, Roy proposed a ReLayNet for fluid segmentation in macular OCT image [26]. Devalla designed the DRUNET for optic nerve head tissue segmentation in OCT image [27]. Venhuizen et al. implement retinal thickness measurement and intraretinal cystoid fluid quantification using the FCN framework [28].

These FCN based segmentation frameworks have achieved promising results. However, most studies utilize a pixel-wise loss, such as softmax, in the last layer for the network, which is insufficient to learn both local and global contextual relations between pixels [29]. To address this problem, researchers have proposed several methods to refine the FCN output and ensure topological relationships of the segmentation results. For example, Ganaye et al. proposed the NonAdjLoss, a loss constraint that suppresses known-forbidden region adjacencies to improve the network's region-labeling consistency in anatomical segmentations [30]. Kepp et al. present an automatical segmentation approach based on shape regression, which employs the signed distance maps to implement spatial regularization and achieved plausible results in retinal OCT image segmentation [31]. He et al. used a combination of two U-nets to segment retinal layers in OCT images. The first U-net segments the several layers whereas the second one refines possible errors in the prediction, thus generating strict topologically correct segmentations [32]. Similar strategy is also presented in Wang's research, which introduces a post processing network to enforce the topology correctness [24]. The topological correction ability of these methods relies on specifically designed cost functions [30,31] or additional post-processing networks [24,32]. The newly designed cost function is generally not easy to be applied for other tasks and additional post-processing structures requires more computational resource.

In 2014, a novel deep learning framework called generative adversarial networks (GAN) [33] was proposed for image generation, which attracts extensive attention from researchers since its performances showed great advantages over the state-of-the-arts [34,35]. Recently, GAN has also been applied to generate segmentation mask in Isola's strategy called Pix2Pix [36]. Thanks to this original work and the following up Pix2PixHD [37], a series of conditional GAN based strategies are introduced to medical image segmentation and generates several attractive researches. Chen et al. used adversarial learning for connectomes segmentation on electron microscopy (EM) images [38]. Liu et al. proposed a semi-supervised method for the

Biomedical Optics EXPRESS

segmentation of layer and fluid region in retinal OCT images using adversarial learning [39]. Tennakoon et al. also proposed a GAN based method for retinal fluid segmentation, which was ranked fourth in the ReTOUCH challenge [40]. Li et al. employed GAN to construct a transfer-learning framework for HEp-2 Specimen Image segmentation [41]. These architectures based on GAN utilize adversarial learning to encode relationships between image pixels, thus eliminating the need for additional post-processing steps. However, segmentation is implemented by the generator, which is supposed to generate "real" images rather than classifying. In this case, it may generate undesirable parts in the label map. Although additional constraints such as $L_1$ norm can alleviate such problems, the accuracy is not so satisfactory as pixel classification.

In this paper, we proposed a novel adversarial convolutional network (ACN), which adopts adversarial learning to train a fully convolutional pixel-wise classifier. The architecture consists of a generator and a discriminator. The generator takes an input image and generates a label map close to the ground truth. The discriminator, which takes the original image and a label map as the input, was trained to identify whether the input pair is real or synthetic. An additional branch is added to the discriminator to implement pixel-wise classification. The architecture of the generator and discriminator are almost the same which is inspired by U-net [25]. Following the proposed framework, the classifier was trained by the adversary of generator and discriminator, which indicates the classification ability was increasingly boosted during the process. The main contributions of this paper can be summarized as follows:

- We describe a novel extension to GANs that enables them to train a U-net alike network for OCT image segmentation.
- A novel architecture is designed for both the generator and discriminator which can be applied to the segmentation task for esophageal OCT images.
- The proposed framework improves segmentation performance on esophageal OCT images with no requirement of additional post-processing.

The rest of this study is organized as follows. Section 2 describes the detailed framework and architecture of the proposed ACN. Section 3 presents experimental settings and segmentation results of ACN on esophageal OCT images, including dataset description, comparison results with widely used deep models and the potential clinical application for EoE diagnosis. Discussions and conclusions are given in Sections 4 and 5, respectively.

## 2. Methods

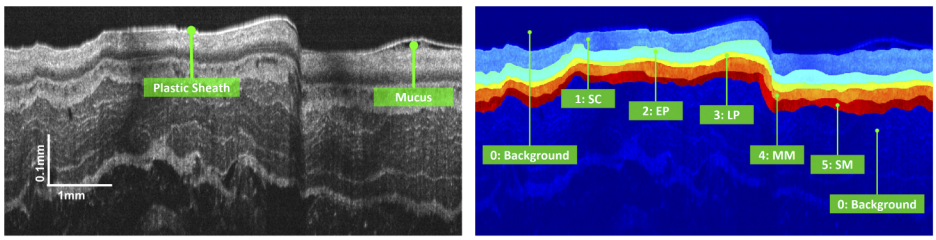
### 2.1. Problem statement

Given an esophageal OCT image $I$, the task is to assign each pixel to a particular label representing a certain tissue. The algorithm proposed in this study will be verified on esophageal OCT images from guinea pigs. A typical image is shown in Fig. 1(a). The tissue layers marked in the images are the epithelium stratum corneum (SC), epithelium (EP), lamina propria (LP), muscularis mucosae (MM) and submucosa (SM), labeled "1" to "5", respectively. The remaining part of the image is treated as the clinically irrelevant region and labeled by "0" as displayed in Fig. 1(b).

### 2.2. ACN framework

Figure 2 introduces the framework of ACN. In this figure, the "real pair" consists of an OCT image slice and the ground truth label map, while the "fake pair" includes an image slice and a label map generated by the network. The validity map is a matrix with the same size as the input image slice with values ranging from 0 to 1 indicating if the input is a "real pair". Ideally, the validity map should be all ones if the input pair is real. The one hot label is a label encoding method by which the categorical variable is converted into a vector that with a single

Biomedical Optics EXPRESS




(a) (b)

**Fig. 1.** Demonstration of (a) a typical esophageal OCT image for the guinea pig and (b) the corresponding manual segmentation result.

"1" and all the others "0". For instance, if a pixel is from the tissue layer labeled by "2", the corresponding one hot label is "[0, 1, 0, 0, 0, 0]". The ACN framework contains two primary components called generator and discriminator. The generator of ACN is designed to obtain a fake label map that is close to the ground truth. Then, the discriminator takes the ground truth label maps or the generated masks along with the original OCT images as input. It is trained to discriminate the synthetic labels from the ground truth and predicts the label for each pixel in the meanwhile. Detailed structures of the generator and discriminator will be explained in the following subsection.

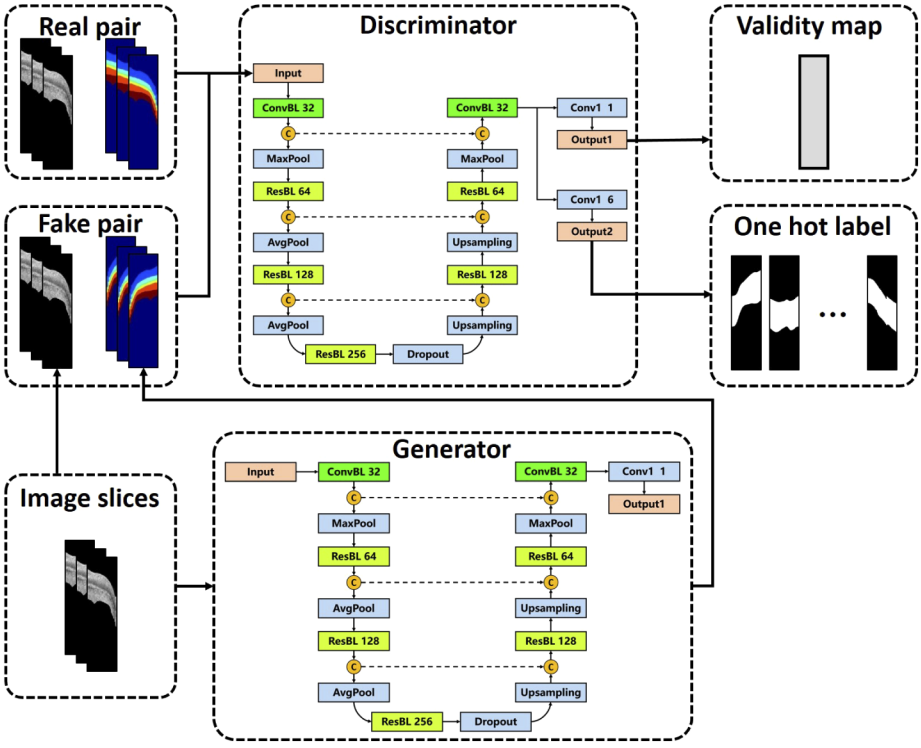



**Fig. 2.** The ACN framework.

As mentioned in the literature [42], in adversarial learning, the generator is always trying to fool the discriminator, which improves the performance of the discriminator and makes it more powerful. Since the generator would produce numerous different outputs, the discriminator will

Biomedical Optics EXPRESS

be trained by images in different conditions even if the real training pair is limited. Moreover, an adversarial network can encode high order relationships between image pixels [40], hence eliminating the need for an additional post-processing step.

### 2.3. Architecture of the generator and discriminator

The generator is designed on the basis of U-Net. Its architecture is presented in Fig. 3, where "ConvBL" indicates the convolution block, "ResBL" means the residual block, "C" represents the concatenate layer, "Conv1" means a $1 \times 1$ convolution layer used for controlling output channels. The number in the block means the output channel. For example, "ConvBL 32" means this convolutional layer generates a 32 channel output. As seen in Fig. 3, the generator architecture has an encoder-decoder structure including downsampling path and upsampling path. The downsampling path gradually reduces spatial resolution and increase the semantic dimension, thus representing the original image by an abstracted feature map. The upsampling path restores the feature map to an image of the same size as the input. In this case, the output is a mask with pixel-wise labels for the input OCT image. The concatenate layer is employed to merge the information from the encoder and decoder, which is intended to fuse features from different scales.

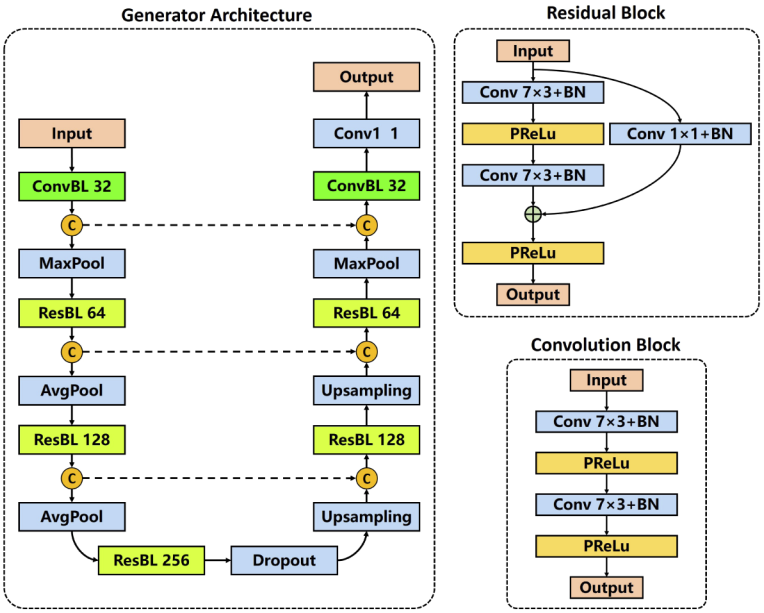



**Fig. 3.** Architecture of the generator.

The generator consists of two specifically designed blocks, namely the convolution block and the residual block shown in Fig. 3. The convolution block is visualized in Fig. 3, which includes two convolution layers, each is followed by a batch normalization layer [43] and a PReLu activation layer [44]. The kernel size is set to $7 \times 3$ to ensure the network to focus more on the intensity variation along the vertical direction [26]. The batch normalization layer is used to compensate for the covariate shifts and is beneficial for a successful training [43,44]. PReLu activation is chosen because it can introduce non-linearity in the training and prevent gradient vanishment. Besides, the PReLu converges faster than ReLu [44].

The residual block was inspired by the Resnet structure [45]. By providing a shortcut connection to transpose the input directly to the output, architecture with residual block has an

Biomedical Optics EXPRESS

effect equivalent to automatically adjust layer numbers. In this case, the residual structure is able to accelerate the convergence of deep networks and improve the classification performance [45]. The residual block for the generator is shown in Fig. 3. The convolution kernel size is set as $7 \times 3$ for the same reason as the convolution block.

The architecture of the discriminator is almost the same as the generator. Differences lie in the output layer as visualized in Fig. 2. The first output generates a single-channel validity map with the same size as the input image. The sigmoid activation is employed and a larger output means that pixels are more likely to come from real pairs. This structure is different for the validation map used by Liu et al. [39] which is optimized by the spatial cross-entropy loss and represents the similarity of the ground truth label and the output of the segmentation network. We use such a validity map since it is easy to be achieved in a full convolutional approach, and the similar format with the segmentation output also makes it easy to be optimized simultaneously. The second output is a six-channel softmax classification result, which assigns each pixel to a certain tissue layer. In this case, the first output measures if the output of generator is real and the second output provides the final segmentation result. Leveraging on this multi-task strategy, the discriminator can discriminate the generated mask and segment the input image at the same time. The trainable parameters of the entire ACN framework are about 42 million.

### 2.4. Loss function

The overall loss function of the ACN can be expressed as Eq. (1).

$$L_{\mathrm{ACN}} = L_{\mathrm{cGAN}}(G, D_1) + \lambda_1 L_{l_1}(G) + \lambda_2 L_{\mathrm{class}}(G, D_2) + \lambda_3 L_{\mathrm{dice}}(G, D_2) \tag{1}$$

In Eq. (1), $G$ is the generator, $D_1$ indicates the validity output of the discriminator and $D_2$ is the class label output of the discriminator. $\lambda_i$ $(i = 1, 2, 3)$ are hyperparameters. In this study, we set them as $\lambda_1 = 100$, $\lambda_2 = 10$, $\lambda_3 = 1$. This loss function is composed of four parts. $L_{\mathrm{cGAN}}$ is the objective of conditional GAN (cGAN) involved in Pix2Pix [36], which is formulated as Eq. (2),

$$L_{\mathrm{cGAN}}(G, D_1) = E_{x,y\sim p_{\mathrm{data}}(x,y)}[\log D_1(x, y)] + E_{x\sim p_{\mathrm{data}}(x)}[\log(1 - D_1(x, G(x)))] \tag{2}$$

where $x$ is the provided conditional image, $y$ is the ground truth label map. $L_{l_1}$ is the $l_1$ distance loss between the synthesized image and the corresponding ground truth for the generator.

$$L_{l_1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1] \tag{3}$$

$L_{\mathrm{class}}$ is a measurement for classification performance, which is generally defined as Eq. (4).

$$L_{\mathrm{class}} = E_{x,y\sim p_{\mathrm{data}}(x,y)}[\log P(C = c \mid x, y)] + E_{x\sim p_{\mathrm{data}}(x)}[\log P(C = c \mid x, G(x))] \tag{4}$$

This study uses the multi-class cross entropy, Eq. (4) can be further expressed as Eq. (5),

$$L_{\mathrm{class}} = -\frac{1}{N}\sum_{i=1}^{N} g_l(f_i) \log p_l(f_i \mid x, y) - \frac{1}{N}\sum_{i=1}^{N} g_l(x_i) \log p_l(f_i \mid x, G(x)) \tag{5}$$

where $N$ is the pixel number, $g_l(f_i)$ is the target probability that pixel $f_i$ belongs to class $l$ with one for the true label and zero entries for the others. $p_l(f_i)$ is the estimated probability of pixel $f_i$ belongs to class $l$. $p_l(f)$ is obtained from the discriminator as described in Eq. (6).

$$\begin{aligned} p_l(f \mid x, y) &= D_2(x, y) \\ p_l(f \mid x, G(x)) &= D_2(x, G(x)) \end{aligned} \tag{6}$$

Biomedical Optics EXPRESS

$L_{\text{dice}}$ is the dice loss aiming at evaluating the spatial overlap of the predicted label and the ground truth, which is defined by Eq. (7),

$$\begin{aligned} L_{\text{dice}} = & \left[1 - \frac{2\sum_{i=1}^{N} p_l(f_i \mid x, y) g_l(f_i)}{\sum_{i=1}^{N} p_l^2(f_i \mid x, y) + \sum_{i=1}^{N} g_l^2(f_i)}\right] \\ & + \left[1 - \frac{2\sum_{i=1}^{N} p_l(f_i \mid x, G(x)) g_l(f_i)}{\sum_{i=1}^{N} p_l^2(f_i \mid x, G(x)) + \sum_{i=1}^{N} g_l^2(f_i)}\right] \end{aligned} \tag{7}$$

where the parameters is defined in the same way as Eq. (5).

Based on the $L_{\text{ACN}}$ defined in Eq. (1), the objective generator and discriminator can be obtained by the optimization defined in Eq. (8).

$$\begin{aligned} G^{\star}, D^{\star} = & \arg(\min_{G}\max_{D}(L_{cGAN}(G, D_1) + \lambda_1 L_{L_1}(G)) \\ & + \min_{G}\min_{D}(\lambda_2 L_{\text{class}}(G, D_2) + \lambda_3 L_{\text{dice}}(G, D_2))) \end{aligned} \tag{8}$$

### 2.5. Training

To get the optimum $G^{\star}$ and $D^{\star}$, this study solves Eq. (8) following a typical strategy which optimizes $G$ and $D$ alternatively. In each iteration, we first train a $D$ with $G$ fixed and then optimize $G$ using the obtained $D$. The optimization is implemented by the Adam method [46] with a learning rate $2 \times 10^{-3}$. Training is performed in batches of 40 randomly chosen samples at each iteration (selected to saturate the GPU memory). After going through the entire training set, an epoch is finished. When finishing 100 training epochs, the model with the lowest validation loss is employed to measure the segmentation performance of the testing dataset for further quantitative evaluation.

The ACN is trained using OCT image slices as shown in Fig. 2. Contrarily, in the testing process, the new-coming image can be sent directly into the network, which benefits from the size-free property of the fully convolutional network, thus obtaining a segmentation result without any slicing induced artifacts [26]. Besides, data augmentation is employed in the training process to overcome the sparsity of training dataset [25] and improve the network robustness to deal with the imbalance in the data set between healthy and sick individuals. The data augmentation techniques in this study include random rotation, horizontal flipping, random shearing, elastic deformations [27].

## 3. Experiments

### 3.1. Data

In this study, 1100 OCT B-scans from guinea pig esophagus were used to evaluate the proposed segmentation networks. These images were collected from different subjects using an 800 nm ultrahigh resolution (axial resolution $\leq 2.5$ $\mu$m) endoscopic OCT system, [47–49], including five healthy samples and two EoE models [4]. As listed in Table 1, these images are divided into three parts. The training set and validation set are used for the development of ACN, which consists of 700 OCT B-scans from four healthy subjects and one EoE subject. An independent dataset with 400 B-scans was collected for testing, which is imaged on another one healthy subject and one EoE subject to ensure no overlaps between training and testing.

Each B-scan from our dataset is of size $2048 \times 2048$ and is resized to $1024 \times 1024$ with a scale factor of 0.5. Considering the fact that the target tissue area exists in the upper half of the image, we crop each B-scan along depth to the size of $512 \times 1024$, which is able to cover all anatomical information. For the data used for training and validation, each B-scan is split width-wise into 8 non-overlapped slices sizing $512 \times 128$. Since our fully-convolutional network can process images of arbitrary size, images in the testing set can be segmented without slicing.

Biomedical Optics EXPRESS

**Table 1. Information of the dataset used in this study.**

| Dataset | Traing set | | Validation set | | Testing set | |
|---|---|---|---|---|---|---|
| Condition | Healthy | EoE | Healthy | EoE | Healthy | EoE |
| Images (Volumes × Frames ) | 5 × 80 | 2 × 80 | 5 × 20 | 2 × 20 | 2 × 100 | 2 × 100 |

The annotated labels were generated by two experienced graders using ITK-SNAP [50], which were used for network training and algorithm evaluation. The ACN was implemented in Keras using Tensorflow as the backend. Training of the network was performed on a 12 GB Tesla K80 GPU using CUDA 9.2 with cuDNN v7.

### 3.2. Evaluation metrics

We use the following metrics to evaluate the proposed ACN framework, including the pixel-wise accuracy (PWA), the dice similarity coefficient (DSC), the average symmetric surface distance (ASSD) and the Hausdorff distance (HD). The PWA and DSC evaluated the segmentation performance based on the overlap area, which are defined as Eq. (9),

$$\begin{aligned} \mathrm{PWA}(A,B) &= \frac{|A \cap B|}{|A|} \\ \mathrm{DSC}(A,B) &= 2 \times \frac{|A \cap B|}{|A| + |B|} \end{aligned} \tag{9}$$

where $A$ and $B$ represent the surface of the ground truth label and the segmentation result, respectively. The ASSD and HD are used to measure the boundary accuracy of the segmentation result as formulated by Eq. (10),

$$\begin{aligned} \mathrm{ASSD}(A,B) &= \frac{1}{2} \times \left[ \frac{\sum_{a \in A} \min_{b \in B} d(a,b)}{|A|} + \frac{\sum_{b \in B} \min_{a \in A} d(b,a)}{|B|} \right] \\ \mathrm{HD}(A,B) &= \max\{\max_{a \in A} \min_{b \in B} d(a,b), \max_{b \in B} \min_{a \in A} d(a,b)\} \end{aligned} \tag{10}$$

where $d(a,b)$ indicates the Eular distance of points $a$ and $b$.

### 3.3. Comparisons with state-of-the-art

We compared the proposed ACN with several state-of-the-art methods in image segmentation, which include the Segnet [51], U-Net [45] and Pix2Pix [36]. Typical segmentation results of different methods for a normal OCT B-scan sample and an EoE one were shown in Figs. 4 and 5.

Raw B-scans and the corresponding label maps for the healthy esophagus are shown in Figs. 4(a) and 4(b). It can be found that the layer structure is transparent and has a uniform thickness. In the Segnet result (Fig. 4(c)), the layer structure is clearly identified and errors occur on the background where certain pixels are considered as tissue. Such error occurs because the pixel classification strategy of Segnet cannot guarantee strict topological relationships. U-Net performs better than the Segnet as shown in Fig. 4(d), because the concatenate structure merged the information from different scales, resulting in a more powerful classification ability. However, U-Net utilized the same pixel classification strategy as Segnet. As a result, topology errors still exist on the SC layer, where some tissues are treated as background, and some pixels are classified as tissues from the SM layer. Segmentation performance of Pix2Pix is demonstrated in Fig. 4(e). Unlike Segnet and U-Net, Pix2Pix segments OCT images by image transforming to generate a label map on condition of provided images. Leveraging on adversarial learning that encodes high order pixel relationships, Pix2Pix results show much fewer topological errors. Figure 4(f) shows the segmentation result of the proposed ACN. This framework combines the advantages of U-Net and Adversarial learning, thus labeling the tissues without layer corruption.

Biomedical Optics EXPRESS

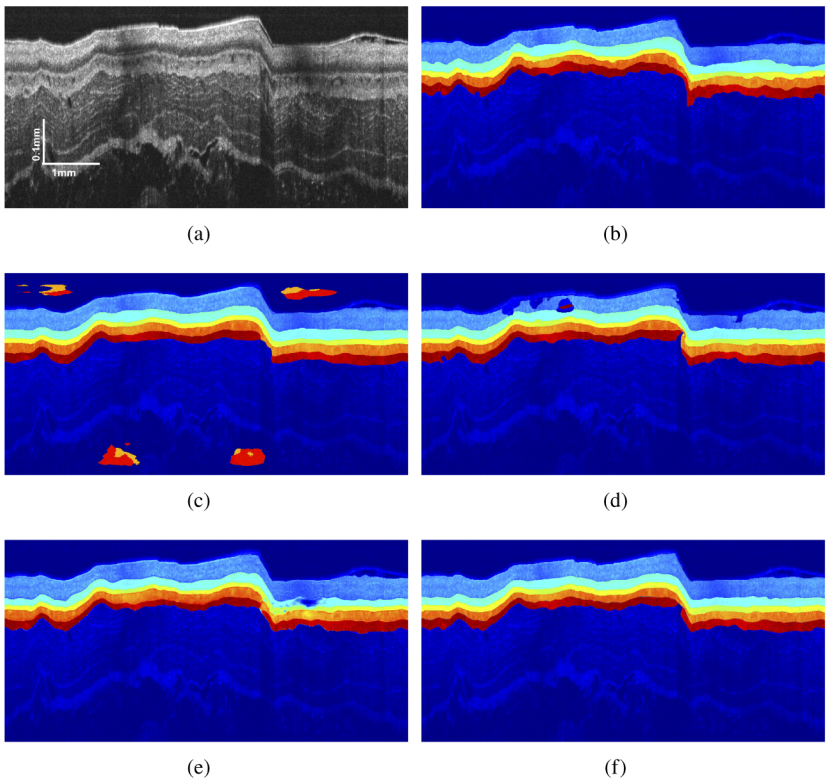



**Fig. 4.** Demonstration of (a) a normal OCT B-scan sample from guinea pig esophagus; (b) manual segmentation ; (c) result of segnet; (d) result of U-Net; (e) result of Pix2Pix and (f) result of the proposed ACN.

For the EoE case, the original B-scan is presented in Fig. 5(a) and the corresponding label is shown in Fig. 5(b). Affected by mucus, the plastic sheath used for protecting the probe cannot stick to the esophageal wall, leading to a large curvature in part of the tissue. Similar to the healthy case, segmentation results of Segnet (Fig. 5(c)) and U-Net (Fig. 5(d)) still suffer from topological errors. The Pix2Pix (Fig. 5(e)) framework alleviates this problem and achieves more complete tissues, but the classification accuracy is not improved. The proposed ACN framework still achieves the highest classification accuracy with no obvious topological errors (Fig. 5(f)). The overall performance of these deep networks in EoE tissue segmentation is inferior to that in the normal case, which indicates the automatical segmentation of diseased esophageal OCT images with mucus and irregular tissues is more challenging.

A more comprehensive evaluation is implemented on the testing dataset consisting of 400 B-scans (200 healthy and 200 EoE) as described in Table 2. In addition to the four deep learning based methods, the table also lists the segmentation result of two graph theory based methods, namely the GTDP [10] and the SBGS [12]. Moreover, manual segmentation result is also presented in the table, where Grader #2 indicates the manual segmentation result of another grader and Grader #1' indicates a second annotation of the same dataset from Grader #1. The automatical segmentation method with the best performance is bolded in the table.

It can be found that the deep learning based methods have higher PWA and DSC that the graph theory based methods, indicating they can identify tissue regions more accurately. Moreover, the deep learning based methods also present smaller ASSD and HD, meaning they also generate tissue surface with fewer errors. Results of U-Net performs better than that of Segnet, confirming

Biomedical Optics EXPRESS

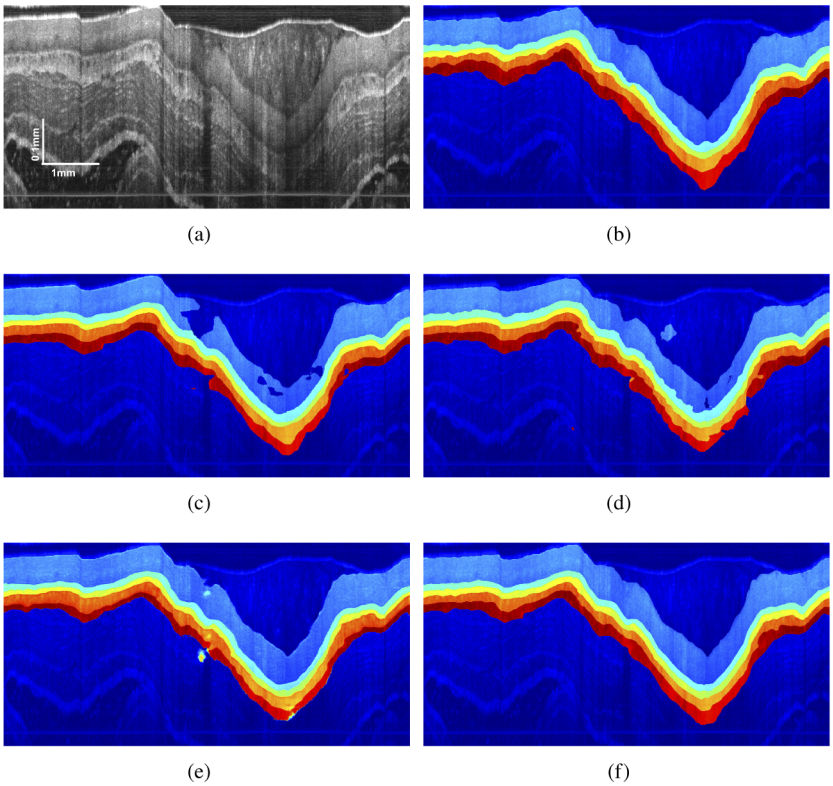



**Fig. 5.** Demonstration of (a) an EoE OCT B-scan sample from guinea pig esophagus; (b) manual segmentation ; (c) result of FCN ; (d) result of U-Net; (e) result of Pix2Pix and (f) result of the proposed ACN.

**Table 2. Metrics of different segmentation methods on esophageal layer segmentation using the annotation from Grader #1 as ground truth.**

| Methods | PWA (%) | DSC (%) | ASSD ($\mu m$) | HD ($\mu m$) |
|---|---|---|---|---|
| Deep Learning Based Methods | | | | |
| Segnet | 95.55 ± 3.17 | 93.67 ± 2.43 | 6.21 ± 2.44 | 42.28 ± 11.67 |
| U-Net | 96.27 ± 3.09 | 94.45 ± 2.37 | 5.87 ± 2.29 | 40.06 ± 10.42 |
| Pix2Pix | 94.69 ± 2.98 | 94.62 ± 2.72 | 5.74 ± 2.15 | 35.53 ± 9.16 |
| ACN | **97.17 ± 2.56** | **95.33 ± 2.18** | **4.57 ± 1.99** | **28.82 ± 8.24** |
| Graph Theory Based Methods | | | | |
| GTDP | 84.45 ± 6.17 | 81.67 ± 7.43 | 10.34 ± 5.44 | 50.89 ± 16.70 |
| SBGS | 86..63 ± 5.04 | 83.55 ± 6.25 | 8.63 ± 4.17 | 47.82 ± 14.33 |
| Manual Segmentation | | | | |
| Grader #2 | 96.01 ± 2.74 | 94.24 ± 2.42 | 5.66 ± 2.48 | 22.18 ± 6.27 |
| Grader #1' | 97.88 ± 2.17 | 95.92 ± 2.14 | 3.25 ± 1.11 | 8.74 ± 6.21 |

Biomedical Optics EXPRESS

the advantages of the concatenate structure. Moreover, U-Net also achieves more accurate segmentation than Pix2Pix, indicating pixel classification is more precise though the tissue layer identified by Pix2Pix seems more reasonable visually. The proposed ACN framework achieves the best performance in all automatical cases, which implies the advantages of the combination of pixel classification and adversarial learning.

The last two rows of Table 2 present the annotation accuracies of graders. The segmentation results of Grader #2 performs similar to the deep learning based methods with an accuracy around 96%. Errors mainly come from some subjective reasons due to the different image interpretation of graders. Besides, the manual segmentation result from the same grader also presents variability in the same dataset since manual annotation may be affected by the working environment and the grader's own conditions. Results proved that the proposed automatical segmentation method can achieve results comparable to manual segmentation.

To evaluate the segmentation performance for each layer individually, the DSCs of different methods for five tissue layers are calculated and listed in Table 3. The SC layer acquires the highest DSCs in all the cases, which is not only because it has relatively large areas, but also results from the fact that this layer is adjacent to the probe thus generating clearer boundaries. For similar reasons, the EP layer is also segmented with high DSCs though it is not of larger area than the other three layers. The proposed ACN framework performs best when segmenting the first three layers among all the tested automatical segmentation algorithms, which confirms its advantages in this task.

**Table 3. DSCs of different segmentation methods for five tissue layers.**

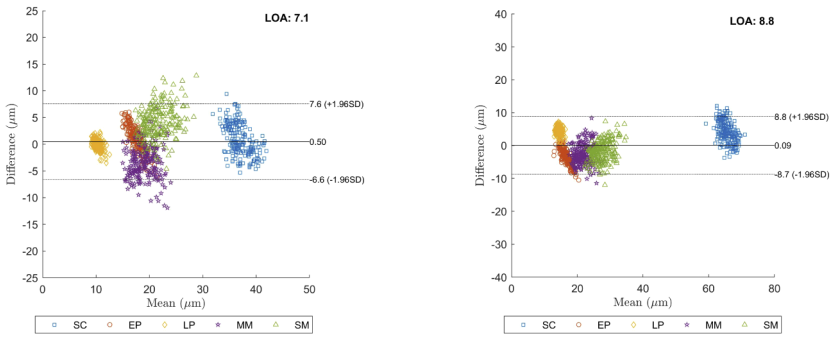
| Methods | SC (%) | EP (%) | LP ($\mu m$) | MM ($\mu m$) | SM |
|---|---|---|---|---|---|
| Deep Learning Based Methods | | | | | |
| Segnet | 97.84 ± 1.51 | 94.76 ± 1.56 | 85.96 ± 5.10 | **82.64 ± 3.07** | 83.99 ± 4.15 |
| U-Net | 97.97 ± 1.37 | 95.05 ± 1.88 | 84.21 ± 4.81 | 82.00 ± 3.56 | **86.79 ± 3.13** |
| Pix2Pix | 97.65 ± 1.56 | 92.30 ± 2.97 | 78.05 ± 2.84 | 77.02 ± 3.57 | 79.34 ± 4.07 |
| ACN | **98.86 ± 1.01** | **96.47 ± 1.82** | **86.47 ± 4.12** | 81.17 ± 3.01 | 83.01 ± 4.29 |
| Graph Theory Based Methods | | | | | |
| GTDP | 86.67 ± 2.28 | 80.18 ± 4.80 | 78.10 ± 4.40 | 73.96 ± 5.43 | 77.36 ± 4.69 |
| SBGS | 88.85 ± 1.76 | 83.86 ± 3.15 | 81.71 ± 3.16 | 79.46 ± 3.99 | 79.95 ± 4.39 |
| Manual Segmentation | | | | | |
| Grader #2 | 97.56 ± 1.66 | 94.81 ± 1.73 | 84.44 ± 4.60 | 82.32 ± 3.51 | 82.01 ± 3.97 |
| Grader #1' | 98.76 ± 1.03 | 96.92 ± 1.75 | 87.25 ± 3.81 | 83.68 ± 2.99 | 85.42 ± 3.67 |

Figure 6 shows the Bland-Altman plot indicating the reliability of the thickness measurements using the proposed ACN algorithm in comparison with the reference annotations from Grader #1. In Fig. 6, LOA represents the limit of agreement with the 95% confidence interval. It can be found that the ACN result generates differences from the annotation labels with around 8 $\mu m$ within a 95 % confidence interval.

### 3.4. Potential applications of ACN in EoE diagnosis

Automatic diagnosis of esophageal diseases relies on characterizing features such as tissue thickness and shape, which can be obtained from the segmentation result. As an important type, EoE is often featured with increased basal zone thickness [4] (Fig. 5). In this case, we calculated the layer thicknesses of five esophageal tissues based on the segmentation output of ACN. Results obtained by ACN and manual segmentation for the testing set are shown in Fig. 7. As was seen from this plot, both the manual and automatical results show that the SC layer for the EoE cases is thicker than the healthy cases, thereby confirming that the layer thickness change

Biomedical Optics EXPRESS




(a) (b)

**Fig. 6.** Bland-Altman plot of the proposed ACN framework compared to the segmentation result from Grader #1 for testing images from (a) the normal case and (b) the EoE case.

is an indicator for EoE. In addition, it is clearly observed that the esophageal segmentation of ACN is consistent with the manual segmentation. For all layers, ACN results show a smaller standard deviation, implying the automatical segmentation works more robustly. Obtaining diagnosis-assistant features from ACN is convenient and accurate, which indicates it is of great potential for practical clinical applications.

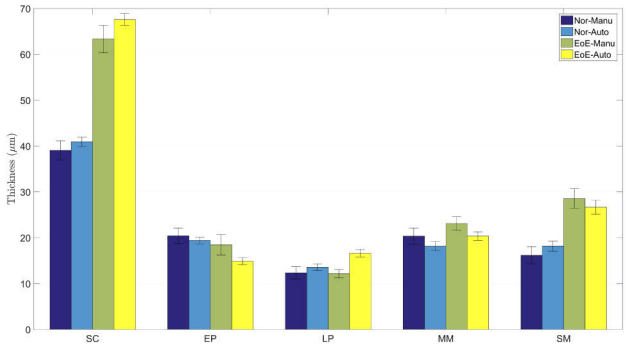



**Fig. 7.** Statistical results of layer thicknesses for guinea pigs of different health conditions, where Auto indicates the result of ACN and Manu represents the thickness achieved from the human grading.

## 4. Discussions

Developing an automatical segmentation system for esophageal OCT images is challenging for numbers of reasons, such as high variability in the appearance of pathology on images, speckle noise and motion artifacts inherent in OCT images. Nowadays, deep learning has become the primary approach in OCT image segmentation since it does not require handcrafted features, and learn features independently on the basis of the training data, thus avoid considering those mentioned problems. However, existing deep frameworks have some limitations adopted directly in esophageal OCT image processing. Our experiments showed that the FCN based methods adopting pixel-wise loss are insufficient to learn topological relationships. As a result, Segnet and U-Net generate ill-posed labels as shown in Figs. 5(c) and 5(d). The GAN based methods like Pix2Pix is able to encode high order relationships between image pixels, thus generating

Biomedical Optics EXPRESS

more continuous label maps with fewer topological errors. However, the segmentation accuracy is not satisfactory comparing to pixel classification as listed in Table 2.

Our framework uses adversarial learning to train a fully convolutional network. In this case, the discriminator is simultaneously optimized by a hybrid loss function including multi-class cross-entropy, GAN loss, $L_1$ loss and dice loss, enabling the network to learn topological relationship and achieve high classification accuracy. As a result, the proposed ACN performs human-like labeling robustly and precisely. Experiment results in Figs. 4(f) and 5(f) confirmed these advantages of ACN. Moreover, extracting diagnostic features using ACN is also convenient. In the experiments, we calculated the SC layer thickness as an indicator for EoE, which showed an evident difference between healthy and diseased esophagus, thus demonstrating the potential of ACN in further clinical applications.

The proposed ACN can be further improved. Firstly, hyperparameter optimization [52] can be adopted to acquire a more accurate classification network. As shown in Eq. (1), the loss function has three hyperparameters to control the weight of different aspects. In this study, these hyperparameters are set experimentally. Hyperparameter optimization can generate more reasonable weights for the network. However, the improvement may not be that evident considering the computation cost. Secondly, more complex networks or some newly developed architectures can also be included in ACN. For example, the U-shape convolutional network in ACN can be changed by more complex semantic segmentation frameworks like DeepLapv3+ [53] and ICNet [54]. The GAN loss can be replaced by loss functions used in LSGAN [55] or WGAN [56]. It is not easy to demonstrate which one is the best, but the idea that segmenting images in an adversarial way is able to boost the original performance.

In the current study, the experiments were based on OCT images with layered esophageal structures from guinea pigs. Esophagus from human shares the same structures as guinea pigs with five tissue layers to be segmented. As a result, the same procedure can be transferred directly to processing esophageal OCT images from human subjects. In the future, endoscopic images collected from other esophageal disease models or human subjects will be studied to improve the proposed method.

## 5. Conclusions

In this study, we introduce the ACN, which uses adversarial learning to train a convolutional network for esophageal OCT image segmentation. The proposed framework takes advantage of pixel classification and adversarial learning, thus generating human-like segmentation results. Experiments on segmenting OCT images from guinea pig esophagus demonstrated that the proposed ACN outperforms the widely used deep learning framework including Segnet, U-Net and Pix2Pix. In addition, the ACN is also able to delineate the OCT images from EoE guinea pig models, which confirmed its potential ability in esophageal disease diagnosis. The proposed ACN introduces a new image segmentation strategy, and its application in esophageal OCT images may facilitate the application of OCT techniques in esophageal disease detection. ACN is convenient for further improvements, such as performing hyperparameter optimization or adding newly developed structures. It is also easy to be transferred for other tasks, such as segmenting esophageal OCT images from other disease models or human subjects. These properties make it appealing for applications in clinical.

## Funding

National Key R&D Program for Major International Joint Research of China (2016YFE0107700); Jiangsu Planned Projects for Postdoctoral Research Funds of China (2018K007A, 2018K044C).

Biomedical Optics EXPRESS

## Acknowledgments

We would like to acknowledge Prof. Xingde Li and Dr. Wu Yuan from the Johns Hopkins University for their technical support in the SD-OCT systems.

## Disclosures

The authors declare that there are no conflicts of interest related to this article.

## References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," Science **254**(5035), 1178–1181 (1991).
2. G. J. Tearney, M. E. Brezinski, B. E. Bouma, S. A. Boppart, C. Pitris, J. F. Southern, and J. G. Fujimoto, "In vivo endoscopic optical biopsy with optical coherence tomography," Science **276**(5321), 2037–2039 (1997).
3. J. M. Poneros, S. Brand, B. E. Bouma, G. J. Tearney, C. C. Compton, and N. S. Nishioka, "Diagnosis of specialized intestinal metaplasia by optical coherence tomography," Gastroenterology **120**(1), 7–12 (2001).
4. Z. Y. Liu, J. F. Xi, M. Tse, A. C. Myers, X. D. Li, P. J. Pasricha, and S. Y. Yu, "Allergic inflammation-induced structural and functional changes in esophageal epithelium in a guinea pig model of eosinophilic esophagitis," Gastroenterology **146**(5), S92 (2014).
5. M. J. Suter, M. J. Gora, G. Y. Lauwers, T. Arnason, J. Sauk, K. A. Gallagher, L. Kava, K. M. Tan, A. R. Soomro, T. P. Gallagher, J. A. Gardecki, B. E. Bouma, M. Rosenberg, N. S. Nishioka, and G. J. Tearney, "Esophageal-guided biopsy with volumetric laser endomicroscopy and laser cautery marking: a pilot clinical study," Gastrointest. Endosc. **79**(6), 886–896 (2014).
6. X. Qi, M. V. Sivak, G. Isenberg, J. E. Willis, and A. M. Rollins, "Computer-aided diagnosis of dysplasia in barrett's esophagus using endoscopic optical coherence tomography," J. Biomed. Opt. **11**(4), 044010 (2006).
7. X. Qi, Y. S. Pan, M. V. Sivak, J. E. Willis, G. Isenberg, and A. M. Rollins, "Image analysis for classification of dysplasia in barrett's esophagus using endoscopic optical coherence tomography," Biomed. Opt. Express **1**(3), 825–847 (2010).
8. D. W. Li, J. M. Wu, Y. F. He, X. W. Yao, W. Yuan, D. F. Chen, H. C. Park, S. Y. Yu, J. L. Prince, and X. D. Li, "Parallel deep neural networks for endoscopic oct image segmentation," Biomed. Opt. Express **10**(3), 1126–1135 (2019).
9. G. J. Ughi, M. J. Gora, A. F. Swager, A. Soomro, C. Grant, A. Tiernan, M. Rosenberg, J. S. Sauk, N. S. Nishioka, and G. J. Tearney, "Automated segmentation and characterization of esophageal wall in vivo by tethered capsule optical coherence tomography endomicroscopy," Biomed. Opt. Express **7**(2), 409–419 (2016).
10. J. L. Zhang, W. Yuan, W. X. Liang, S. Y. Yu, Y. M. Liang, Z. Y. Xu, Y. X. Wei, and X. D. Li, "Automatic and robust segmentation of endoscopic oct images and optical staining," Biomed. Opt. Express **8**(5), 2697–2708 (2017).
11. M. Gan, C. Wang, T. Yang, N. Yang, M. Zhang, W. Yuan, X. D. Li, and L. R. Wang, "Robust layer segmentation of esophageal oct images based on graph search using edge-enhanced weights," Biomed. Opt. Express **9**(9), 4481–4495 (2018).
12. C. Wang, M. Gan, N. Yang, T. Yang, M. Zhang, S. H. Nao, J. Zhu, H. Y. Ge, and L. R. Wang, "Fast esophageal layer segmentation in oct images of guinea pigs based on sparse bayesian classification and graph search," Biomed. Opt. Express **10**(2), 978–994 (2019).
13. L. Y. Fang, N. J. He, S. T. Li, P. Ghamisi, and J. A. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," IEEE Trans. Geosci. Remote Sensing **56**(3), 1803–1815 (2018).
14. L. Y. Fang, C. Wang, S. T. Li, H. Rabbani, X. D. Chen, and Z. M. Liu, "Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification," IEEE Trans. Med. Imaging **38**(8), 1959–1970 (2019).
15. R. Rasti, M. J. Allingham, P. S. Mettu, S. Kavusi, K. Govind, S. W. Cousins, and S. Farsiu, "Deep learning-based single-shot prediction of differential effects of anti-vegf treatment in patients with diabetic macular edema," Biomed. Opt. Express **11**(2), 1139–1152 (2020).
16. Z. Y. Han, B. Z. Wei, A. Mercado, S. Leung, and S. Li, "Spine-gan: Semantic segmentation of multiple spinal structures," Med. Image Anal. **50**, 23–35 (2018).
17. Y. G. Shi, K. Cheng, and Z. W. Liu, "Hippocampal subfields segmentation in brain mr images using generative adversarial networks," BioMed. Eng. OnLine **18**(1), 5 (2019).
18. D. Romo-Bucheli, P. Seebock, J. I. Orlando, B. S. Gerendas, S. M. Waldstein, U. Schmidt-Erfurth, and H. Bogunovic, "Reducing image variability across oct devices with unsupervised unpaired learning for improved segmentation of retina," Biomed. Opt. Express **11**(1), 346–363 (2020).
19. J. Wang, T. T. Hormel, L. Q. Gao, P. X. Zang, Y. K. Guo, X. G. Wang, S. T. Bailey, and Y. L. Jia, "Automated diagnosis and segmentation of choroidal neovascularization in oct angiography using deep learning," Biomed. Opt. Express **11**(2), 927–944 (2020).

Biomedical Optics EXPRESS


20. H. Stegmann, R. M. Werkmeister, M. Pfister, G. Garhofer, L. Schmetterer, and V. A. Dos Santos, "Deep learning segmentation for optical coherence tomography measurements of the lower tear meniscus," Biomed. Opt. Express **11**(3), 1539–1554 (2020).
21. L. Y. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. T. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search," Biomed. Opt. Express **8**(5), 2732–2744 (2017).
22. J. Kugelman, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Automatic segmentation of oct retinal boundaries using recurrent neural networks and graph search," Biomed. Opt. Express **9**(11), 5759–5777 (2018).
23. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *2015 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)* pp. 3431–3440 (2015).
24. J. Wang, Z. Wang, F. Li, G. X. Qu, Y. Qiao, H. R. Lv, and X. L. Zhang, "Joint retina segmentation and classification for early glaucoma diagnosis," Biomed. Opt. Express **10**(5), 2639–2656 (2019).
25. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Med. Image Comput. Comput. Interv. Pt III **9351**, 234–241 (2015).
26. A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," Biomed. Opt. Express **8**(8), 3627–3642 (2017).
27. S. K. Devalla, P. K. Renukanand, B. K. Sreedhar, G. Subramanian, L. Zhang, S. Perera, J. M. Mari, K. S. Chin, T. A. Tun, N. G. Strouthidis, T. Aung, A. H. Thiery, and M. J. A. Girard, "Drunet: a dilated-residual u-net deep learning network to segment optic nerve head tissues in optical coherence tomography images," Biomed. Opt. Express **9**(7), 3244–3265 (2018).
28. F. G. Venhuizen, B. van Ginneken, B. Liefers, F. van Asten, V. Schreur, S. Fauser, C. Hoyng, T. Theelen, and C. I. Sanchez, "Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography," Biomed. Opt. Express **9**(4), 1545–1569 (2018).
29. Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. L. Huang, "Segan: Adversarial network with multi-scale l (1) loss for medical image segmentation," Neuroinform. **16**(3-4), 383–392 (2018).
30. P. A. Ganaye, M. Sdika, B. Triggs, and H. Benoit-Cattin, "Removing segmentation inconsistencies with semi-supervised non-adjacency constraint," Med. Image Anal. **58**, 101551 (2019).
31. T. Kepp, J. Ehrhardt, M. P. Heinrich, G. Huttmann, and H. Handels, "Topology-preserving shape-based regression of retinal layers in oct image data using convolutional neural networks," in *2019 Ieee 16th International Symposium on Biomedical Imaging (Isbi 2019)*, (2019), pp. 1437–1440.
32. Y. F. He, A. Carass, Y. H. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Deep learning based topology guaranteed surface and mme segmentation of multiple sclerosis subjects from retinal oct," Biomed. Opt. Express **10**(10), 5042–5058 (2019).
33. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27 (Nips 2014)*, vol. 27 (2014).
34. D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," IEEE Trans. Biomed. Eng. **65**(12), 2720–2730 (2018).
35. F. Mahmood, R. Chen, and N. J. Durr, "Unsupervised reverse domain adaptation for synthetic medical images via adversarial training," IEEE Trans. Med. Imaging **37**(12), 2572–2581 (2018).
36. P. Isola, J. Y. Zhu, T. H. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *30th Ieee Conference on Computer Vision and Pattern Recognition (Cvpr 2017)*, (2017), pp. 5967–5976.
37. T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *2018 Ieee/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr)*, (2018), pp. 8798–8807.
38. K. Chen, D. D. Zhu, J. W. Lu, and Y. Luo, "An adversarial and densely dilated network for connectomes segmentation," Symmetry **10**(10), 467 (2018).
39. X. M. Liu, J. Cao, T. Y. Fu, Z. F. Pan, W. Hu, K. Zhang, and J. Liu, "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," IEEE Access **7**, 3046–3061 (2019).
40. R. Tennakoon, A. K. Gostar, R. Hoseinnezhad, and A. Bab-Hadiashar, "Retinal fluid segmentation in oct images using adversarial loss based convolutional neural networks," *2018 Ieee 15th International Symposium on Biomedical Imaging (Isbi 2018)* pp. 1436–1440, (2018).
41. Y. X. Li and L. L. Shen, "cc-gan: A robust transfer-learning framework for hep-2 specimen image segmentation," IEEE Access **6**, 14048–14058 (2018).
42. A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," (2016).
43. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," (2015).
44. K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 Ieee International Conference on Computer Vision (Iccv)*, (2015), pp. 1026–1034.
45. K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)*, (2016), pp. 770–778.
46. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," (2014).

Biomedical Optics EXPRESS


47. J. F. Xi, A. Q. Zhang, Z. Y. Liu, W. X. Liang, L. Y. Lin, S. Y. Yu, and X. D. Li, "Diffractive catheter for ultrahigh-resolution spectral-domain volumetric oct imaging," Opt. Lett. **39**(7), 2016–2019 (2014).
48. W. Yuan, J. Mavadia-Shukla, J. F. Xi, W. X. Liang, X. Y. Yu, S. Y. Yu, and X. D. Li, "Optimal operational conditions for supercontinuum-based ultrahigh-resolution endoscopic oct imaging," Opt. Lett. **41**(2), 250–253 (2016).
49. W. Yuan, R. Brown, W. Mitzner, L. Yarmus, and X. D. Li, "Super-achromatic monolithic microprobe for ultrahigh-resolution endoscopic optical coherence tomography at 800 nm," Nat. Commun. **8**(1), 1531 (2017).
50. P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," NeuroImage **31**(3), 1116–1128 (2006).
51. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," (2015).
52. L. A. Thiede and U. Parlitz, "Gradient based hyperparameter optimization in echo state networks," Neural Netw. **115**, 23–29 (2019).
53. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," (2018).
54. H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "Icnet for real-time semantic segmentation on high-resolution images," (2017).
55. X. D. Mao, Q. Li, H. R. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *2017 Ieee International Conference on Computer Vision (Iccv)*, (2017), pp. 2813–2821.
56. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," http://arxiv.org/abs/1701.07875 (2017).